Cell PRESS
Structure
Article

Structure and Function of the *E. coli* Dihydroneopterin Triphosphate Pyrophosphatase: A Nudix Enzyme Involved in Folate Biosynthesis

Sandra B. Gabelli,[1,3] Mario A. Bianchet,[1,3] WenLian Xu,[2] Christopher A. Dunn,[2] Zhi-Dian Niu,[2] L. Mario Amzel,[1,4,*] and Maurice J. Bessman[2,4]
[1] Department of Biophysics and Biophysical Chemistry, School of Medicine, Johns Hopkins University, 725 North Wolfe Street, Baltimore, MD 21205, USA
[2] Department of Biology and McCollum-Pratt Institute, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
[3] These authors contributed equally to this work.
[4] These authors contributed equally to this work.
*Correspondence: mario@neruda.med.jhmi.edu
DOI 10.1016/j.str.2007.06.018

SUMMARY

Nudix hydrolases are a superfamily of pyrophosphatases, most of which are involved in clearing the cell of potentially deleterious metabolites and in preventing the accumulation of metabolic intermediates. We determined that the product of the *orf17* gene of *Escherichia coli*, a Nudix NTP hydrolase, catalyzes the hydrolytic release of pyrophosphate from dihydroneopterin triphosphate, the committed step of folate synthesis in bacteria. That this dihydroneopterin hydrolase (DHNTPase) is indeed a key enzyme in the folate pathway was confirmed in vivo: knockout of this gene in *E. coli* leads to a marked reduction in folate synthesis that is completely restored by a plasmid carrying the gene. We also determined the crystal structure of this enzyme using data to 1.8 Å resolution and studied the kinetics of the reaction. These results provide insight into the structural bases for catalysis and substrate specificity in this enzyme and allow the definition of the dihydroneopterin triphosphate pyrophosphatase family of Nudix enzymes.

INTRODUCTION

Nudix hydrolases (Bessman et al., 1996) are a superfamily of pyrophosphatases catalyzing the hydrolysis of a nucleoside diphosphate linked to any of several moieties. Some of the substrates are metabolites that require modulation during the cell cycle or during periods of stress. Others are deleterious compounds that must be eliminated to protect cell integrity. Most of the Nudix hydrolases are "house-cleaning enzymes" that control the levels of these compounds. A few Nudix hydrolases have been shown to catalyze intermediate steps in specific biosynthetic pathways (Espinosa et al., 1999; Klaus et al., 2005). Across species, the members of the superfamily contain the signature sequence $GX_5EX_7REUXEEXG/TU$ (U = L, V, I) (Bessman et al., 1996; O'Handley et al., 1998). This sequence, the Nudix motif, forms the catalytic and metal binding sites of more than 4000 enzymes in species ranging from viruses to prokaryotes to eukaryotes.

Folate biosynthesis in bacteria has been extensively studied. The enzymes involved in this pathway in *Escherichia coli* and their genes have been identified, with the exception of the enzyme that catalyzes dihydroneopterin triphosphate (DHNTP) hydrolysis to dihydroneopterin monophosphate (DHNMP) and pyrophosphate (Bermingham and Derrick, 2002) (Figure S1; see the Supplemental Data available with this article online). As this reaction resembles the hydrolysis of nucleotide triphosphates by Nudix hydrolases, we suspected that the *E. coli* DHNTPase was a Nudix enzyme. In 2005, Hanson and coworkers identified a DHNTP pyrophosphatase in *Lactococcus lactis* but did not identify its counterpart in *E. coli* (Klaus et al., 2005). Thirty years earlier, in 1974, Suzuki and Brown described such an activity in *E. coli* and proposed that, in combination with other phosphatases, this putative enzyme yielded dihydroneopterin, the next intermediate in the synthesis of folate. The protein they purified was a 17 kDa enzyme with a K_m of 11 μM and depended on Mg^{2+} for activity (Suzuki and Brown, 1974). However, the protein was not sequenced and the *E. coli* gene coding for the enzyme was not identified. Looking for the *E. coli* DHNTPase, we examined the previously (O'Handley et al., 1996) characterized *E. coli* Nudix enzymes and found that one enzyme, the product of the *orf17* gene (nudB, ntpA), had characteristics similar to those of the enzyme described by Suzuki and Brown (1974).

The product of the *orf17* gene had been shown to catalyze the hydrolysis of dATP to dAMP and pyrophosphate (dNTPase) (O'Handley et al., 1996) with a maximum activity at pH 8.5 in the presence of 5 mM Mg^{2+}. However, in this paper, we show that this enzyme is the *E. coli* DHNTP pyrophosphatase (DHNTPase) and that a knockout of this

Table 1. Kinetic Data, dATP versus DHNTP

Substrate	V_{max} (Units/mg)	K_m (mM)	k_{cat} (s^{-1})	k_{cat}/K_m (M/s^{-1})
dATP	15.0 ± 0.73	0.79 ± 0.073	4.59	5,810
DHNTP	37.9 ± 1.93	0.27 ± 0.037	11.6	43,000

V_{max} and K_m were determined from Lineweaver-Burk plots (Lineweaver and Burk, 1934) using a nonlinear least-squares fit weighted to substrate concentration (Cleland, 1979).

gene has highly impaired folate synthesis. Thus, this work identifies the long sought after enzyme that catalyzes the second step in the synthesis of folate in *E. coli*. Except for the Nudix signature sequence, this enzyme has no sequence similarity to the DHNTPase from *L. lactis* (Klaus et al., 2005).

To fully characterize this member of a Nudix family, we determined the structure of the enzyme by X-ray diffraction (1.8 Å) and identified residues involved in substrate specificity and in catalysis. Combination of the three-dimensional structure and sequence data provides key information about the structural bases of substrate specificity in this enzyme and allows the identification of other members of a possible Nudix DHNTPase family. Thus, this work presents the basis for identifying the enzyme that catalyzes this activity in other species by searching the Nudix sequences for the presence of the residues that recognize the substrate.

RESULTS

Comparison of the Kinetic Constants for the Hydrolysis of dATP and DHNTP by the Enzyme

Our earlier characterization of Orf17 indicated that it catalyzed the hydrolysis of all the canonical ribo- and deoxyribonucleoside triphosphates, with a preference for dATP (O'Handley et al., 1996). Table 1 shows that the enzyme has a higher k_{cat} and a lower K_m for DHNTP than for dATP, such that the catalytic efficiency (k_{cat}/K_m) for DHNTP is over seven times greater than for dATP. V_{max} of this enzyme with DHNTP (38 μmol/min/mg of protein) is approximately 60- and 40-fold faster than that reported for the *Lactococcus* and *Arabidopsis* enzymes, respectively (Klaus et al., 2005), and over 2000-fold greater than that reported by Suzuki and Brown (1974).

In Vivo Function of the dATPase/DHNTPase

Whether DHNTPase is crucial for folic acid production in *E. coli* was tested by deleting the *orf17* gene. Although there was no reduction in growth rate in the knockout compared to parental strains in either M9 minimal medium (supplemented with proline and leucine) or in Luria Broth (Sambrook and Russell, 2001), the folic acid concentration in the *orf17*$^-$ strain was markedly reduced (Figure 1). When the *orf17* gene was introduced into the deleted strain in a plasmid, folic acid production was completely restored (Figure 1). The residual concentration of folic acid in the knockout strain could result from the nonenzymatic conversion of DHNTP to the monophosphate (De Saizieu et al., 1995) and/or from the action of other nonspecific pyrophosphatases in the cell. Noteworthy is the small but significant increase in folic acid in the two strains supplemented with the plasmid carrying the *orf17* gene (bars b and d) compared to parent cells supplemented with the plasmid lacking *orf17* (bar a). This could be related to the observations of Klaus et al. (2005), who measured the concentration of pteridine derivatives in wild-type *L. lactis* and found that DHNTP was ten times higher than the corresponding monophosphate, suggesting that DHNTPase hydrolysis is rate limiting. The increase in folic acid in the cells containing extra copies of *orf17* could thus result from an overproduction of DHNTPase.

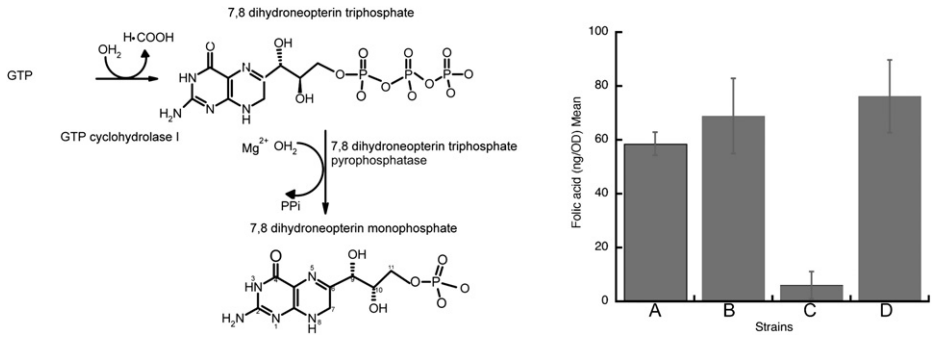

Figure 1. Scheme of the 7,8 Dihydroneopterin Triphosphate Pyrophosphatase (DNHTPase) Catalytic Reaction and Effect of the *orf17* Deletion on *E. coli*'s Folate Concentration

(Left) Reaction catalyzed by the product of the gene *orf17*, DHNTPase. (Right) Folic acid concentration of parental cells and cells made devoid of DHNTPase. Each bar represents the mean of four separate cultures, two of which were grown in M9 minimal medium and two in Luria Broth. Cells were harvested and assayed microbiologically using *Lactobacillus casei* (ATTC 7469) (Wilson and Horne, 1982). Bar a: parent cells containing plasmid without insert (MG1655 harboring pTrc 99A); bar b: parent cells harboring plasmid with *orf17* (MG1655-pTrc 99A-Orf17); bar c: parent with deleted *orf17* harboring plasmid without insert (MG1655*orf17*$^-$ – pTrc 99A); bar d: parent with deleted *orf17* harboring plasmid with *orf17* insert (MG1655*orf17*$^-$ – pTrc 99A-Orf17).

Table 2. Data Collection and Refinement Statistics of DHNTPase Hydrolase

Crystal	EMTS1	EMTS2	SMCL	Native 1 (Room Temperature)	Native 2 (Refinement)
Space group	C2				
Cell dimensions	a = 124.1 Å, b = 43.2 Å, c = 108.0 Å, β = 115.0°				a = 124.1 Å
					b = 42.6 Å
					c = 106.4 Å
					β = 115.6°
Data Collection					
X-ray source	Raxis IIC (room temperature)				BNL-X4A (100K)
Wavelength (Å)	CuKa 1.541 Å				1.1 Å
Resolution (high-resolution shell) (Å)	50.0–2.7 (2.8–2.7)	50.0–2.7 (2.8–2.7)	19.97–2.6 (2.69–2.6)	65.0–2.05 (2.1–2.05)	50.0–1.8 (1.98–1.80)
Measured reflections	31,743	34,048	35,920	239,801	302,686
Unique reflections	13,091	13,548	14,304	33,047	46,996
Completeness (%)	89.7 (86.9)	99.6 (99.9)	87.7 (81.0)	96.2 (78.2)	95.6 (95.7)
R_{merge} (%)	13.1 (31.0)	11.4 (38.0)	11.5 (32.2)	8.7 (21.6)	8.6 (34.0)
Refinement					
R_{cryst} (%)			20.2 (25.5)	19.2 (21.0)	22.9 (25.6)
R_{free}			28.0 (36.8)	27.9 (29.0)	28.1 (33.8)
Rms deviations					
Bond length (Å)			0.009	—	0.009
Angle (°)			1.2	—	1.2
Monomer in asymmetric unit			4	—	4
Total atoms			4,818	—	5,364
Protein atoms			4,706	—	4,706
Water molecules			69	—	611
Ligand (per monomer)			2 SO_4 (A, B)		2 SO_4 (A, B)
			1 PPi (A, C)		1 PPi (A, B, C)
Metals (per monomer)	13 Hg	7 Hg	Sm^{3+} (A, B, C) 1 Na (A, C)	—	—
B factor (protein) ($Å^2$)	—	—	34.9	—	20.2
B factor (ligand) ($Å^2$)	—	—	67.0		45.1
B factor (H_2O) ($Å^2$)	—	—	26.6	—	33.3

Structure of *E. coli* DHNTPase

To identify which residues are involved in substrate specificity and in catalysis, we determined the structure of *E. coli* DHNTPase. The structure was determined by MIRAS and refined with data to 1.8 Å resolution to a final R_{crys}/R_{free} of 0.22/0.28 (Table 2). The atomic model (four monomers A, B, C, and D in the asymmetric unit) has excellent geometry, with the majority of the residues (93.8%) located in the most favored region of the Ramachandran plot and the remaining 6.2% located in the allowed region. Interactions among the four molecules in the asymmetric unit do not involve large interfaces (less than 900 $Å^2$ between A and B; about 280 $Å^2$ between A and C; and 500 $Å^2$ between A and D), in agreement with the observation that *E. coli* DHNTPase is a monomer in solution (O'Handley et al., 1996).

Each monomer is formed by a simple Nudix fold (Gabelli et al., 2001), in which loop L5, joining strands β5 and β6 of the mixed β sheet, has a short insertion of a 3_{10} helix (Figure 2A). The twisted mixed β sheet, packed by helices on both sides, forms the walls of the active site where the substrate binds. Whereas monomer D has an empty binding site, monomers A, B, and C have a pyrophosphate bound. Monomers A (residues A2–A148), B (residues B4–B149), and C (residues C6–C147), probably as a result of having pyrophosphate bound, are observed in the same

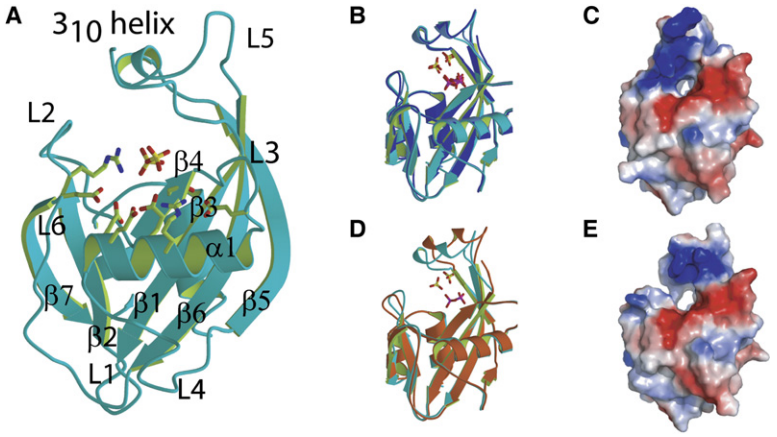

Figure 2. Structure of the *E. coli* Dihydroneopterin Pyrophosphatase

(A) Ribbon diagram of the structure of DHNTPase in complex with its product, pyrophosphate. Residues of the Nudix signature sequence, Arg-29 and Glu-117, are shown as green sticks. The secondary structural elements are labeled L for loops, β for strands, and α for helices.

(B) Ribbon diagram of the overlap of monomers A, B, and C (turquoise, cornflower blue, and blue, respectively) with their pyrophosphate and sulfate molecules.

(C) Molecular surface of the complex structure (monomer A, closed).

(D) Ribbon diagram of the overlap of monomer A (complex structure; turquoise) and monomer D (ligand-free structure, open; orange-red).

(E) Molecular surface of the ligand-free structure (open).

"closed" conformation, with root-mean-square (rms) deviations between 0.33 and 0.36 Å for their equivalent Cα atoms (Figures 2B and 2C). In contrast, monomer D (residues D7–D148) is in a different conformation, with an rms deviation of 1.06 Å with respect to monomer A (Figures 2D and 2E). The positions of loops L2 and L5 and the 3_{10} helix account for the largest differences (Figure 2C). In monomers A, B, and C, the 3_{10} helix (residues 86–97) moves inward by about 5 Å at the Cα position of residue 89 to create the binding site of the pyrophosphate. The "open" conformation observed in monomer D may represent the structure of the free enzyme. In this structure, the first six residues are disordered (Figure 2E).

Nudix Motif

In DHNTPase, a β strand-loop-helix-loop motif (β4-L3-α1-L4; Figure 2) containing the Nudix signature sequence extends from Gly-41 to Ile-63 (Figure 3A), and is stabilized by a network of hydrogen bonds among conserved residues. The C terminus of helix α1 of the Nudix motif is anchored to the mixed β sheet by a hydrogen bond between carboxylate Oε1 of Glu-60 and the main-chain amide of the residue preceding the first glycine of the Nudix motif, Thr-40 (Figure 3A). The carbonyl of Glu-59 is at hydrogen-bonding distance from residue Thr-116 of loop L6, and Thr-62 Oγ is at hydrogen-bonding distance from the carbonyl of Ile-114. Both interactions connect helix α1 of the Nudix motif to loop L6. At the other end of the motif, a salt bridge between the carboxylate of Glu-47 and the guanidinium of Arg-55 clamp the helix to loop L3. The salt bridge and the hydrogen bonds to the carbonyl of Ser-42 and the amide of Glu-44 clamp helix α1 and loop L3 to strand β4 of the mixed sheet (Figure 3A). The side chains of residues Val-57, Val-61, and Ile-63 form a hydrophobic patch on the side of the helix-loop portion of the Nudix motif that rests on the mixed β sheet.

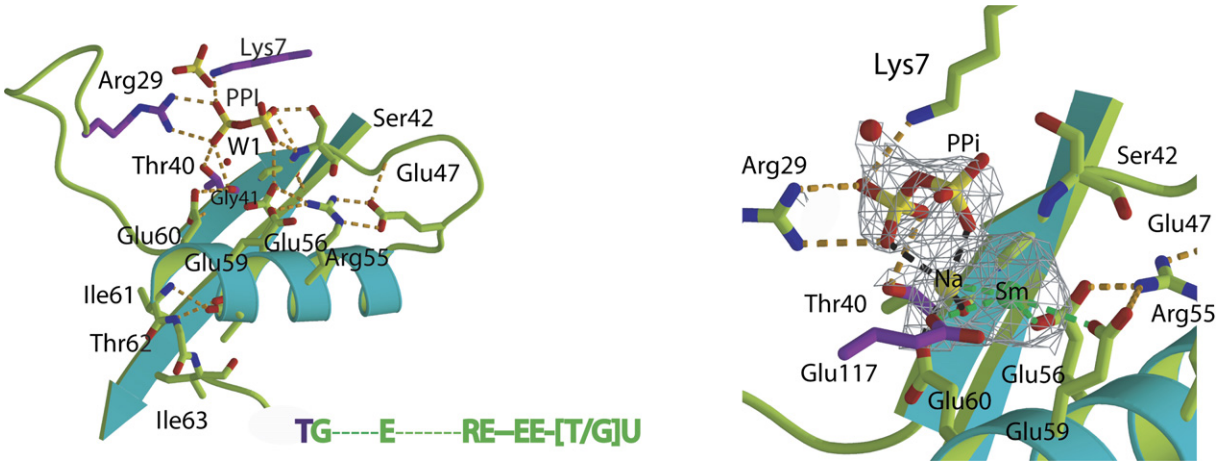

Figure 3. Interactions of the Signature Sequence Nudix Residues of DHNTPase

(Left) Hydrogen-bonding pattern of the conserved residues (orange dashes) and recognition of the pyrophosphate inhibitor as observed in the native complex structure (monomers A, B, and C). (Right) Recognition of the pyrophosphate as observed in the complex in the presence of $SmCl_3$. A $2F_o - F_c$ electron density map of the metal and pyrophosphate is shown contoured at 1.2σ.

Recognition of Pyrophosphate

Pyrophosphate, the product of the hydrolysis of DHNTP and dNTPs, and an inhibitor of the enzyme (Ki = 19 μM) (O'Handley et al., 1996), was added to the crystallization medium to improve crystal quality as in other Nudix enzymes (Gabelli et al., 2001, 2004; Kang et al., 2003b). In the crystal structure of DHNTPase, three of the monomers (A, B, and C) have pyrophosphate in the active site (Figure 2B), where its oxygen atoms make hydrogen bonds to protein residues and a water molecule (Figure 3A). In particular, the α-phosphate (phosphate closest to the first residue of the Nudix motif; see below) has its oxygen atoms at hydrogen-bonding distances from the positively charged residues Lys-7 (Nε) and Arg-29 (Nω2 and Nω1) and from the side-chain hydroxyl and main-chain carbonyl of Thr-40. In monomers A, B, and C, the presence of pyrophosphate and, most likely, substrate, orders Lys-7, as suggested by its change in conformation compared to the structure of the free enzyme (monomer D; Figure 2D). In the other phosphate (β-phosphate), one oxygen atom is at hydrogen-bonding distance from the side-chain Oγ and the main-chain amide of Ser-42 and the Oε2 of Glu-56, a Nudix signature sequence residue. The bridging oxygen of the pyrophosphate is recognized by a water molecule (W1). Another phosphate oxygen is at hydrogen-bonding distance from Arg-29, which in other Nudix complexes coordinates the α-phosphate. Thus, most likely, the pyrophosphate is binding in the position of the α-β phosphates and not the β-γ phosphates of the substrate, suggesting that it is not binding as product but as a substrate/inhibitor (Figures 2B and 2D).

Metal Binding Site: Structure of DHNTPase in Complex with Sm^{3+}

Lanthanides have been frequently found to bind to Mg^{2+} sites (Dudev et al., 2005; Gabelli et al., 2001; Glusker, 1991; Kang et al., 2003a, 2003b). Crystals soaked in samarium chloride, the heavy-atom derivative used for the structure determination (Table 2), reveal unambiguously the samarium binding site of the DHNTPase due to its anomalous signal at the copper edge (data not shown). The samarium binding site, seen in monomers A and C, is formed by the carboxylates of glutamate residues of the Nudix signature sequence Glu-56 (bidentate), Glu-59, and Glu-117. In monomer C, for example, the oxygen metal distances are 2.19 and 2.84 Å, 3.04 and 3.20 Å, respectively, forming a square plane that contains the Sm^{3+}. The position above the plane is occupied by a sodium atom (Figure 3B), whereas the sixth position is empty. Equivalent residues were observed to coordinate a gadolinium ion in the *E. coli* Nudix ADPRase (Gabelli et al., 2001). The position of the samarium atom does not seem to be the appropriate location for the catalytic metal.

The sodium atom is liganded with octahedral coordination, with a square plane formed by one oxygen of each phosphate (Na^+-O distance, 2.24 and 2.81 Å) on one side and the carboxylates of Glu-56 and Glu-60 on the other. The position above the plane is occupied by the carbonyl oxygen of the Thr-40 (distance, 2.51 Å), and the samarium atom occupies the position below at 2.9 Å. This atom is perfectly positioned to act as a catalytic metal. The absence of a side chain at Gly-41 creates a cavity for the catalytic metal, and allows a close approach of the pyrophosphate and Glu-59 to the metal ion. As the sodium atom has been identified solely on the bases of its coordination and interatomic distances, its true chemical identity is unknown.

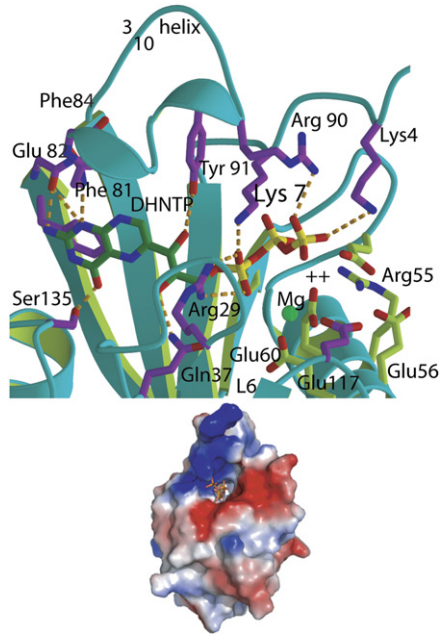

Figure 4. Structural Basis of the Specificity of DHNTPase for DHNTP

(Top) Ribbon diagram of the model of DHNTPase in complex with its substrate dihydroneopterin triphosphate (rotated 90° from Figure 1A) and Mg^{2+}. In green are the residues of the Nudix signature sequence; in purple are other conserved residues believed to be involved in recognition or catalysis. The carbons of DHNTP are shown in green. The green sphere shows the modeled magnesium ion coordinated with ligands of the Nudix motif and Glu-117. Dihydroneopterin is stacked between Phe-81 and Phe-84 (front; not shown for clarity). Orange dashes denote hydrogen bonds. (Bottom) Molecular surface of the closed conformation of DHNTPase, with DHNTP in the cavity shown as sticks.

Structural Basis for Substrate Specificity

Binding of DHNTP and Mg^{2+} to the enzyme was modeled using the observed positions of pyrophosphate and Na^+ in the DHNTPase + PPi + Sm^{3+} complex structure and knowledge of the position of the substrates in other Nudix enzymes (Protein Data Bank ID codes 1KTG and 1JKN).

The modeled DHNTP fits in the cavity observed in the closed structure (Figures 2C and 4B), with the dihydroneopterin stacked between Phe-81 and Phe-84. The carbonyl of the pterin ring is at hydrogen-bonding distance from the Oγ of Ser-135; the substituent amine and the ring nitrogen N1 are at hydrogen-bonding distance from the carbonyl of Glu-82 and the amide of Phe-84 (Figure 4).

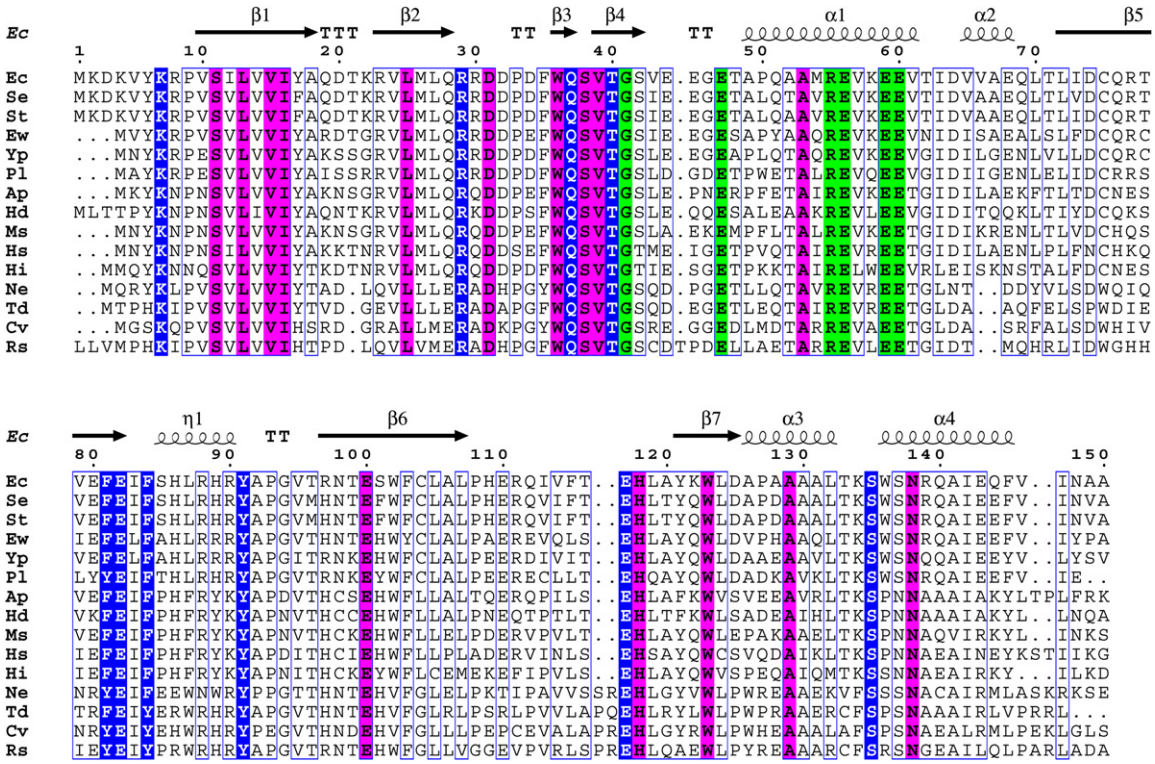

Figure 5. Sequence Alignment of the Proposed DHNTPase Family

Sequence alignment of the proposed DHNTPase family with the secondary structure of the E. coli DHNTPase shown at the top. The residues are colored according to Figure 3: the conserved Nudix signature sequence is shown in green; conserved residues involved in recognition or catalytic mechanism are colored in blue; other conserved residues are shown in magenta. Protein sequences are Ec (NP_288302, *E_coli*_O157), Se (NP_456467, *S_enterica*_Typhi), St (NP_460857, *S. typhimurium*_LT2), Ew (YP_050587, *Erwinia carotovora*), Yp (NP_405606, *Y. pestis*), Pl (NP_929368, *P. luminescens*_TTO1), Ap (YP_001053845, *A. pleuropneumoniae*), Hd (NP_873138, *H. ducreyi* 35000HP), Ms (AAU37316, *M. succinici producens*), Hs (ZP_00132407, *Hs. somnus*_2336), Hi (ZP_00156157, *H. influenzae*_R2866), Ne (NP_842255, *N. europaea*_19718), Td (ZP_00333351, *T. denitrificans*_25), Cv (NP_902630, *C. violaceum* 12472), Rs (NP_518589, *R. solanacearum*_GMI10). This figure was made with the programs ClustalW (Jeanmougin et al., 1998) and ESPript (Gouet et al., 1999).

The hydroxyl groups are at hydrogen-bonding distance of Gln-37 and Tyr-91, and the side chains of Arg-29 and Lys-7 are at hydrogen-bonding distance from the oxygen of the α-phosphate, most likely to stabilize the leaving group of the hydrolysis (dihydroneopterin monophosphate). Arg-90 (either arginine or lysine in the shown DHNTPase sequences; Figure 5) and Lys-4 recognize the γ-phosphate. This cluster of positively charged residues is responsible for product inhibition by pyrophosphate (Figure 4A). Interestingly, two of the residues that recognize the dihydroneopterin ring (Phe-84 and Tyr-91) are in the 3_{10} helix of loop L5, which may participate in the hinge motion that accompanies substrate binding in other Nudix hydrolases.

We modeled the carboxylates of Nudix signature sequence residues Glu-56, Glu-60, and Glu-117 as metal ligands. In Nudix enzymes, the metal that bridges the substrate to the enzyme always includes a glutamate (or a glutamine) residue from loop L6 as a ligand. The position modeled for the Mg^{2+} coincides with the position of the Na atom in the crystal structure of the samarium complex.

Identification of Other Members of the DHNTPase Family

A sequence search of the *Arabidopsis* genome with the previously described *L. lactis* DHNTPase sequence identified an analogous enzyme (At1g6876) that had significant DHNTPase activity. However, attempts to align the sequence of the *L. lactis* DHNTPase (Ll-DHNTPase) with that of the *E. coli* enzyme (Ec-DHNTPase) showed no significant homology outside the region of the Nudix signature. In contrast, when we carried out a search of the nonredundant (NCBI, NIH) database with the Ec-DHNTPase sequence, we identified 15 bacterial sequences with significant homology outside the Nudix box (Figure 5). We suggest that these proteins are all members of the DHNTPase family.

The residues postulated to be involved in dihydroneopterin recognition, Lys-7, Arg-29, Gln-37, Thr-40, Phe-81, Glu-82, Phe-84, Tyr-91, and Ser-135, together with the fourth ligand to the catalytic metal, Glu-117, are conserved in the family of Nudix DHNTPase enzymes (Figure 5). Sequence alignments of DHNTPases analyzed in light of the model show that all residues involved in the

substrate recognition are conserved in the DHNTPase family (Figures 4A and 5).

The identification of these 15 homologs of the Ec-DHNTPase suggests that this family is widely used in bacterial folate biosynthesis. In contrast, the Ll-DHNTPase, with few identified analogs, represents a less common evolutionary solution to the catalysis of DHNTP hydrolysis.

Proposed Catalytic Mechanism

The presence of both an open (no pyrophosphate; monomer D) and a closed (pyrophosphate-bound; monomers A, B, and C) conformation argues in favor of a conformational change in which loops L2 and L5 (carrying the 3_{10} helix) close on the active site upon substrate binding. The modeled DHNTP suggests a catalytic mechanism in which residues Glu-82, Phe-84, and Ser-135 recognize the dihydroneopterin and Arg-29 and Lys-7 compensate for the negative charge that develops in the leaving group after hydrolysis. A set of positively charged residues, Arg-30, Arg-90, and Lys-4, makes up a surface that recognizes the phosphates of the triphosphate. There are two possibilities for the catalytic base: His-118 of loop L6 and Tyr-91, located in the helix 3_{10} of loop L5. The argument in favor of His-118 is supported by the observation that in dimeric Nudix hydrolases the base is in loop L6, the loop that closes toward the binding site upon substrate binding. However, in DHNTPase, His-118 is "too far" from the α-phosphate, even in the closed conformation, to play this role. On the other hand, loop L5 of the DHNTPase, which also moves upon substrate binding, contains Tyr-91, a residue that may act as a base. This residue appears to be the most likely catalytic base. Supporting the idea that the conformational changes are due to pyrophosphate binding, and so are involved in the catalytic mechanism, is the analogous conformational change observed in the equivalent helix of the *Lupinus angustifolius* Nudix diadenosine tetraphosphatase upon ATP binding (Fletcher et al., 2002).

DISCUSSION

Nudix hydrolases comprise a superfamily of enzymes that catalyze the hydrolysis of nucleoside diphosphate derivatives and related compounds including nucleoside triphosphates, ADP-ribose, dinucleoside polyphosphates, coenzymes, nucleotide sugars, inositol pyrophosphate, and capped RNA (for reviews, see Bessman et al., 1996; McLennan, 2006). The Nudix hydrolase genes are widely distributed in nature, from viruses to humans. What started as a family of 20 known members in 1996 has now grown into a superfamily of over 4000, as annotated in a recent survey of the data banks. Almost all organisms whose genomes have been sequenced contain multiple representatives of this gene superfamily, ranging in number from 1 in the archaeon *Methanococcus jannaschii* (She et al., 2006) to 30 in *Bacillus anthracis* (Xu et al., 2004). Over 50 Nudix enzymes have been characterized so far. Most appear to be molecules involved in housecleaning functions—removing toxic metabolites or controlling the concentration of metabolic side products that can be deleterious or that carry out regulatory activities. Only two of the enzymes, the product of the *Streptomyces alboniger pur7* gene (Espinosa et al., 1999) and the *L. lactis* dihydroneopterin triphosphatase (Klaus et al., 2005), have been shown to catalyze steps in biosynthetic pathways.

Although not recognized at the time of its discovery, the enzyme that catalyzes the second step in the folic acid biosynthetic pathway in *E. coli* has properties that are distinctly characteristic of the Nudix hydrolase superfamily: it is a relatively small protein (MW 17,000 kDa), requires Mg^{2+}, has a pH optimum of 8.5, and releases PP_i instead of P_i from the triphosphate (Suzuki and Brown, 1974). We synthesized DHNTP, and when we tested it as a substrate of Orf17, an enzyme that we had previously characterized as a dNTPase (O'Handley et al., 1996), we found that it was a highly active DHNTPase. Based on this observation, we concluded that this enzyme catalyzes the second step, the committed step in folate biosynthesis. To test this hypothesis, we investigated the effect of deleting the *orf17* gene, characterized the *orf17* gene product including kinetic studies, determined its three-dimensional structure, and established its relevance in in vivo folic acid synthesis in *E. coli*.

In folate biosynthesis, GTP cyclohydrolase catalyzes the opening of the ribose furanose ring of GTP with loss of one carbon as formate to yield 7,8-dihydroneopterin triphosphate (DHNTP). The enzyme described in this work, DHNTPase, catalyzes the hydrolysis of DHNTP to the monophosphate (DHNMP) plus pyrophosphate. Deletion of the *orf17* gene of *E. coli* yields a bacterium with highly reduced folate levels that can be restored by transformation with a plasmid carrying the *orf17* gene. These data show unambiguously that the gene product of the *E. coli orf17* is indeed a DHNTPase, and that the enzyme is the phosphoanhydrase that catalyzes the second step in folate biosynthesis. This was the only step in folate biosynthesis for which the enzyme responsible had neither been identified nor characterized chemically or structurally. The work presented here fills a gap in our knowledge of folate biosynthesis: it brings the description of the second step in the synthesis from being carried out by an unidentified enzyme or enzymes to a well-defined reaction, catalyzed by a newly identified and characterized protein for which the structure and the mechanism are now known. Furthermore, we identified bacterial homologs of this enzyme that may represent a family of Nudix DHNTP hydrolases.

The enzymatic characterization, combined with amino acid sequence analysis and structure determination, provides insights into the structural basis for substrate specificity in this family of Nudix hydrolases: of all conserved residues in the DHNTPase family, Lys-7, Arg-29, Gln-37, Thr-40, Phe-81, Glu-82, Phe-84, Tyr-91, and Ser-135 are directly involved in substrate specificity and Glu-117 is involved in metal binding. As folate biosynthesis has been established as a target for antibacterial compounds, the structure and the mechanism of DHNTPase described here may provide a basis for drug design.

EXPERIMENTAL PROCEDURES

Enzyme Assays

Measurements of DHNTPase activity were done with enzyme purified and assayed according to O'Handley et al. (1996). The colorimetric assay measures the conversion of an inorganic pyrophosphatase (PP_iase)-insensitive substrate to a PP_iase-sensitive product:

$$\text{X-triphosphate} + H_2O \rightarrow \text{X-monophosphate} + PP_i$$

$$PP_i + H_2O \rightarrow 2P_i,$$

where X is a pterin or a nucleoside. The first reaction occurs with the addition of DHNTPase and the second with PP_iase.

The standard incubation mixture contains, in 50 μl, 2 mM substrate, 50 mM Tris (pH 8.5), 5 mM $MgCl_2$, 0.5 units of yeast inorganic pyrophosphatase, and 0.3–2 milliunits of DHNTPase enzyme (a unit of enzyme catalyzes the hydrolysis of 1 μmol of substrate per min under these conditions).

After 15 min at 37°C, the reaction is terminated by the addition of 50 μl of a mixture of four parts Norit (20% packed volume) and one part 7% perchloric acid, which adsorbs all nonhydrolyzed substrate. After centrifugation, an aliquot of the supernatant is analyzed for inorganic orthophosphate by the method of Ames and Dubin (1960).

Construction of the DHNTPase Knockout Strain

The gene coding for Orf17 was deleted by homologous recombination (Datsenko and Wanner, 2000) from *E. coli* strain MG1655 using upstream and downstream primers encompassing the entire gene. Complete deletion was verified by PCR analysis of the region.

Assay of Folic Acid

The concentration of folic acid in the parental and knockout strains was measured using *Lactobacillus casei* based on the microbiological procedure of Wilson and Horne (1982).

Synthesis of DHNTP

The gene coding for GTP cyclohydrolase (GenBank accession number X63910) catalyzing the conversion of GTP to DHNTP was cloned and expressed in *E. coli*, and the protein, purified by two successive ammonium sulfate fractionations, was used to synthesize DHNTP. The DHNTP was purified on a column of DEAE Sephadex using a triethylammonium bicarbonate eluant, and the salt was removed by successive evaporations and washes with methanol in vacuo.

Crystallization of the DHNTP Pyrophosphatase

Purified DHNTP pyrophosphatase (O'Handley et al., 1996) was stored at −80°C in 50 mM Tris HCl (pH 7.5), 0.1 mM EDTA, 1 mM DTT at concentrations of 11.0–18.0 mg/ml. The enzyme was 95% pure as judged by SDS-PAGE stained with Coomassie blue. The best crystals grew by microseeding in sitting-drop experiments with a 1 ml reservoir solution of 1.3–1.5 M ammonium sulfate, 1% propanol, 3–5 mM DTT, 4 mM sodium pyrophosphate, 100 mM Na HEPES (pH 6.8). The drop was made of one volume of seeds, one volume of reservoir solution, and two volumes of protein.

Data Collection of the DHNTP Pyrophosphatase

DHNTP crystals belong to monoclinic space group C2 with cell dimensions a = 124.1 Å, b = 43.2 Å, c = 108.0 Å, and β = 115.0°. Despite their low solvent content, 26.9% (Mathews coefficient of 1.69), the DHNTP crystals were sensitive to damage during flash-freezing. To overcome the resulting lack of isomorphism among frozen crystals, data for phase determination were collected in house, at room temperature, using a Raxis IIC image plate detector with a copper target rotating anode RU200 X-ray generator as the source. For the final crystal structure refinement, data were collected from a native crystal flash-frozen using 15% glycerol as cryoprotectant (native 2; Table 2) at beamline X4A at the National Synchrotron Light Source, Brookhaven National Laboratory (λ = 1.1 Å) recorded in individual image plates and read with a Fuji scanner. Oscillation images were processed with DENZO and scaled using SCALEPACK (Otwinowski and Minor, 1997). Patterson maps calculated with the structure factors of the native data set showed strong peaks corresponding to a noncrystallographic translation of (0.25, 0.479, 0.0), consistent with a noncrystallographic two-fold axis parallel to the crystallographic b axis intersecting the X,Z plane at X = 0.125, Z = 0.24.

Phase Determination and Structure Refinement

Crystals were soaked overnight in mother-liquor solutions containing either 1 mM ethyl-mercury thiosalicylate (EMTS) or 1 mM samarium (III) chloride (SMCL) (Cate and Doudna, 1996) (Table 2). Mercury heavy-atom sites identified in a difference Patterson map and verified using the program RSPS (Bricogne et al., 2003) were used to find the structures of the other derivatives by inspection of difference Fourier maps. The heavy-atom structures of the three derivative crystals (EMTS1, EMTS2, and SMCL; Table 2) refined together using MLPHARE (CCP4, 1994) were used to estimate new phases. Identification of additional sites and further refinement of the heavy-atom structure carried out with AUTOSHARP (Bricogne et al., 2003) yielded a final overall figure of merit of 0.48 to 2.6 Å resolution. Noncrystallographic symmetry (NCS) operators identified using the programs O, MAPMAN, and MAMA (Kleywegt, 2000; Kleywegt and Jones, 1999) were used to improve the quality of the map by density averaging and phase extension to 2.05 Å with the program DM. The final model, containing four DHNTPase molecules and three pyrophosphates in the crystal asymmetric unit, was refined using a maximum likelihood residual with the program REFMAC. Anisotropic refinement using translation, libration, and screw rotation (TLS) of rigid bodies was carried out using each monomer as a TLS group (Winn et al., 2001).

The full structure of the $SmCl_3$-containing crystals was refined with REFMAC using NCS constraints between monomers A, B, and C. No NCS constraints were applied to monomer D to allow for the different conformations adopted by loop L5-3_{10} helix (residues 82–97).

Modeling of the DHNTPase-Mg^{2+}-DHNTP Complex

A model of the DHNTPase-Mg^{2+}-DHNTP complex was built based on the DHNTPase-PP_i crystal structure, using the molecular modeling program Quanta (Accelrys). DHNTP models were docked manually using the pyrophosphate and sulfate molecules as markers and placing the dihydroneopterin ring in the cavity observed in the molecular surface. The pyrophosphate conformation was adjusted to improve recognition by Arg-29; in all the other steps of the procedure, the positions of the protein atoms remained fixed. Models were locally minimized to decrease the number of unfavorable interactions. The final model shows no unfavorable interactions and maximizes the number of hydrogen-bonding partners.

Supplemental Data

Supplemental Data include one figure and can be found with this article online at http://www.structure.org/cgi/content/full/15/8/1014/DC1/.

ACKNOWLEDGMENTS

Support was provided by NIGMS grants GM066895 (to L.M.A.) and GM18649 (to M.J.B.). Beamlines X25, X4A, and X6A of the National Synchrotron Light Source, Brookhaven National Laboratory are gratefully acknowledged. We thank Dr. J. Navaza and Dr. A.M. Silva for discussions on early attempts to determine the structure by molecular replacement.

Received: April 20, 2007
Revised: June 19, 2007
Accepted: June 22, 2007
Published: August 14, 2007

REFERENCES

Ames, B.N., and Dubin, D.T. (1960). The role of polyamines in the neutralization of bacteriophage deoxyribonucleic acid. J. Biol. Chem. *235*, 769–775.

Bermingham, A., and Derrick, J.P. (2002). The folic acid biosynthesis pathway in bacteria: evaluation of potential for antibacterial drug discovery. Bioessays *24*, 637–648.

Bessman, M.J., Frick, D.N., and O'Handley, S.F. (1996). The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. J. Biol. Chem. *271*, 25059–25062.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr. D Biol. Crystallogr. *59*, 2023–2030.

Cate, J.H., and Doudna, J.A. (1996). Metal-binding sites in the major groove of a large ribozyme domain. Structure *4*, 1221–1229.

CCP4 (Collaborative Computational Project, Number 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Cleland, W.W. (1979). Statistical analysis of enzyme kinetic data. Methods Enzymol. *63*, 103–139.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR product. Proc. Natl. Acad. Sci. USA *97*, 6640–6645.

De Saizieu, A., Vankan, P., and van Loon, A.P. (1995). Enzymic characterization of *Bacillus subtilis* GTP cyclohydrolase I. Evidence for a chemical dephosphorylation of dihydroneopterin triphosphate. Biochem. J. *306*, 371–377.

Dudev, T., Chang, L.Y., and Lin, C. (2005). Factors governing the substitution of La^{3+} for Ca^{2+} and Mg^{2+} in metalloproteins: a DFT/CDM study. J. Am. Chem. Soc. *127*, 4091–4103.

Espinosa, J.C., Tercero, J.A., Rubio, M.A., and Jimenez, A. (1999). The pur7 gene from the puromycin biosynthetic pur cluster of *Streptomyces alboniger* encodes a Nudix hydrolase. J. Bacteriol. *181*, 4914–4918.

Fletcher, J.L., Swabrick, J.D., Maksel, D., Gayler, K.R., and Gooley, P.R. (2002). The structure of Ap4A hydrolase complexed with ATP-MgFx reveals the bases of substrate binding. Structure *10*, 205–213.

Gabelli, S.B., Bianchet, M.A., Bessman, M.J., and Amzel, L.M. (2001). The structure of ADP-ribose pyrophosphatase reveals the structural basis for the versatility of the Nudix family. Nat. Struct. Biol. *8*, 467–472.

Gabelli, S.B., Bianchet, M.A., Azurmendi, H.F., Xia, Z., Saraswat, V., Mildvan, A.S., and Amzel, L.M. (2004). Structure and mechanism of GDP-mannose glycosyl hydrolase, a Nudix enzyme that cleaves at carbon instead of phosphorus. Structure *12*, 927–935.

Glusker, J.P. (1991). Structural aspects of metal liganding to functional groups in proteins. Adv. Protein Chem. *42*, 1–77.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305–308.

Jeanmougin, F., Thompson, J.D., Gouy, M., Higgins, D.G., and Gibson, T.J. (1998). Multiple sequence alignment with Clustal X. Trends Biochem. Sci. *23*, 403–405.

Kang, L.W., Gabelli, S.B., Bianchet, M.A., Xu, W.L., Bessman, M.J., and Amzel, L.M. (2003a). Structure of a coenzyme A pyrophosphatase from *Deinococcus radiodurans*: a member of the Nudix family. J. Bacteriol. *185*, 4110–4118.

Kang, L.W., Gabelli, S.B., Cunningham, J.E., O'Handley, S.F., and Amzel, L.M. (2003b). Structure and mechanism of MT-ADPRase, a Nudix hydrolase from *Mycobacterium tuberculosis*. Structure *11*, 1015–1023.

Klaus, S.M., Wegkamp, A., Sybesma, W., Hugenholtz, J., Gregory, J.F., III, and Hanson, A.D. (2005). A Nudix enzyme removes pyrophosphate from dihydroneopterin triphosphate in the folate synthesis pathway of bacteria and plants. J. Biol. Chem. *280*, 5274–5280.

Kleywegt, G.J. (2000). Structure validation I. In International Tables for Crystallography, M.G. Rossmann and E. Arnold, eds. (Dordrecht, The Netherlands: Kluwer Academic), Chapter 21.21.

Kleywegt, G.J., and Jones, T.A. (1999). Software for handling macromolecular envelopes. Acta Crystallogr. D Biol. Crystallogr. *55*, 941–944.

Lineweaver, H., and Burk, D. (1934). The determination of enzyme dissociation constants. J. Am. Chem. Soc. *56*, 658–666.

McLennan, A.G. (2006). The Nudix hydrolase superfamily. Cell. Mol. Life Sci. *63*, 123–143.

O'Handley, S.F., Frick, D.N., Bullions, L.C., Mildvan, A.S., and Bessman, M.J. (1996). *Escherichia coli* orf17 codes for a nucleoside triphosphate pyrophosphohydrolase member of the MutT family of proteins. Cloning, purification, and characterization of the enzyme. J. Biol. Chem. *271*, 24649–24654.

O'Handley, S.F., Frick, D.N., Dunn, C.A., and Bessman, M.J. (1998). Orf186 represents a new member of the Nudix hydrolases, active on adenosine(5′)triphospho(5′)adenosine, ADP-ribose, and NADH. J. Biol. Chem. *273*, 3192–3197.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. *277*, 307–326.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning, Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

She, M., Decker, C.J., Chen, N., Tumati, S., Parker, R., and Song, H. (2006). Crystal structure and functional analysis of Dcp2p from *Schizosaccharomyces pombe*. Nat. Struct. Mol. Biol. *13*, 63–70.

Suzuki, Y., and Brown, G.M. (1974). The biosynthesis of folic acid. XII. Purification and properties of dihydroneopterin triphosphate pyrophosphohydrolase. J. Biol. Chem. *249*, 2405–2410.

Wilson, S.D., and Horne, W.D. (1982). Use of glycerol-cryoprotected *Lactobacillus casei* for microbiological assay of folic acid. Clin. Chem. *28*, 1198–1200.

Winn, M.D., Isupov, M.N., and Murshudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. Acta Crystallogr. D Biol. Crystallogr. *57*, 122–133.

Xu, W., Dunn, C.A., Jones, C.R., D'Souza, G., and Bessman, M.J. (2004). The 26 Nudix hydrolases of *Bacillus cereus*, a close relative of *Bacillus anthracis*. J. Biol. Chem. *279*, 24861–24865.

Accession Numbers

Coordinates have been deposited in the Protein Data Bank under ID codes 2O1C (DHNTPase in complex with pyrophosphate) and 2O5W (DHNTPase in complex with samarium chloride and pyrophosphate).

